DOCUMENT RESUME

ED 156 707 TM 007 153

TITLE	Assessment: Collection of Papers. Coombe Lodge Report, Vol. 7 No. 13.
INSTITUTION	Further Education Staff Coll., Blagdon (England).
PUB DATE	74
NOTE	35p.
AVAILABLE FROM	Report Editor, Coombe Lodge, Blagdon, Bristol BS18 6RG, England (prices available upon request)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; *Achievement Tests; Course Objectives; Data Processing; *Educational Assessment; Evaluation Criteria; *Evaluation Methods; Evaluation Needs; External Degree Programs; Foreign Countries; *Measurement Goals; National Programs; Objective Tests; *Post Secondary Education; *Student Evaluation; Test Construction
IDENTIFIERS	City and Guilds of London Institute; *Great Britain

ABSTRACT

The five papers included in this collection illustrate the current concerns of those involved in the development and implementation of assessment procedures. A one year in-course assessment procedure conducted at Leicester University is described, including the advantages and disadvantages of the procedure. The problems of interpreting the results of overall assessment grades at open universities and the future possibilities for a procedure involving criterion-based elements, norm-based elements and student skill profiles are discussed in an article on assessment at the Open University of University College, Cardiff. The use of objective tests at the tests at the City and Guilds of London Institute (CGLI) are described. The advantages of objective tests include thorough and uniform syllabus coverage, reduced emphasis on penmanship, reliability marking, statistical analysis, rapid issue of results, and increased opportunity for teacher participation. The author contends that objective tests can be utilized for in-course, entry, and end-of-term evaluations. The use of data processing in assessment at CGLI is outlined including computer analysis and distribution, result analysis, item analysis, and test statistics for objective tests. Selection of criteria and methods for the assessment of students are discussed, with particular reference to the practice of Britain's National Examination Board in Supervisory Studies. (JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED156707

Coombe Lodge Report

Vol 7 No 13

Collection of Papers

Assessment

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

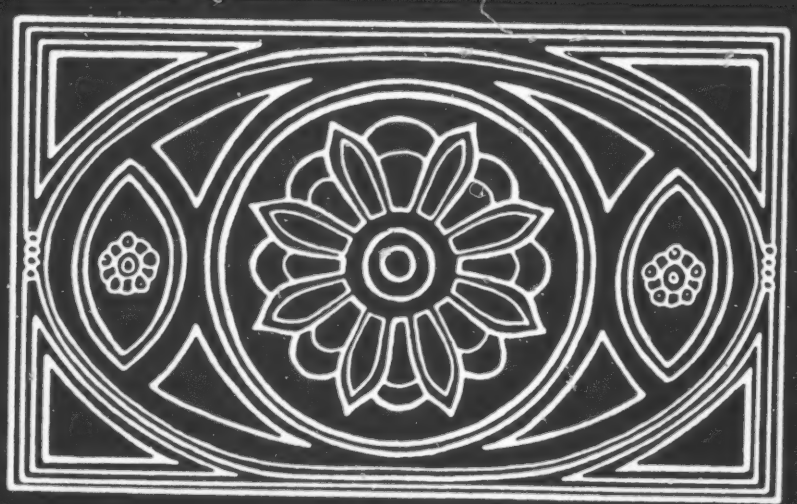
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Jean Finlayson
Report Editor

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM007 153



COOMBE LODGE STAFF

Director:	A H Yates
Deputy Director & Director of Studies:	G E Wheeler
Registrar & Clerk to the Governors:	A F Mitchard
Senior Staff Tutor:	D C Spencer
Staff Tutors:	A Albrecht
	C T Ferguson
	J Latcham
	T J Russell
	C M Turner
Head of Resources Group:	D L Parkes
Research Associate:	P S Hibberd
Librarian:	Christine Ball

COOMBE LODGE REPORTS

Editor:	Philippa Andrews
Secretary:	Suzanne Shipsey

Copyright C The Further Education Staff College, 1974.

Printed offset lithography on Crown quarto by Messrs. Hiscox & Lambert Ltd. Bristol.

ASSESSMENT Vol.7. No.13.
1974.

Collection of Papers

CONTENTS

Foreword 788

In-course Assessment 789
Professor J L Eggleston, Head of Colleges Division, School of Education, University of Nottingham.

Examinations and Assessment at the Open University 795
Dr G D Moss, Director, Centre for Educational Technology, University College, Cardiff.

Developments in CGLI Examination at Craft and Operative Levels 802
Mrs C Ward, Examination Techniques Developments Unit, City and Guilds of London Institute.

Data Processing of City and Guilds Examinations 808
A Taylor, Head of Computer Services, City and Guilds of London Institute.

Identifying Factors for Assessment 810
B Argent, Administrative Officer, City and Guilds of London Institute.

FOREWORD

The five papers included in this collection illustrate the current concerns of various bodies involved in assessment procedures. The value of in-course assessment is considered by Professor Eggleston. Dr Moss describes the various methods used in the validation of Open University students. Two representatives of the City and Guilds of London Institute give an account of recent CGLI developments; Mrs Ward describes developments in assessing craft and operative level students and Mr Taylor outlines the CGLI system of data processing. Finally, Mr Argent considers which factors it is possible to measure in the assessment of students, with particular reference to the practice of the National Examination Board in Supervisory Studies.

Professor J L Eggleston, Head of Colleges Division, School of Education, University of Nottingham

IN-COURSE ASSESSMENT

"I suppose that one of the main trends which can be discerned in the practice of examining in this country is the increasing use made of a variety of assessment procedures. Of these procedures, in-course assessment is the most controversial. I take in-course assessment to be a generic term which includes assessments usually made by teachers of pupils' attainments while a course is still in progress, in contrast to "terminal" assessments made by outside agencies. In considering the relative merits of different assessment procedures, such as practical tests, short questions demanding answers in continuous prose, long questions demanding short answers, objectively marked test and so on, it is necessary to be clear about the purpose which the assessment is designed to serve. The purposes served by assessment procedures are these:

functions of assessment

1. Selection

As long as the number of vacancies at any level in the educational system or in employment is less than the number of candidates aspiring to fill these vacancies, then some sort of evidence will be collected and used for the purpose of selection. This selective function implies that assessment can function predictively.

2. Public evidence attainment

It is frequently claimed that externally imposed assessment procedures provide objective measures of attainment described in a linguistic currency which is accepted and understood. This evidence is claimed to be such that it provides for a necessary accountability of a student's husbandry of his own talents.

3. The carrot and stick hypothesis

It has been argued that students and/or teachers may be so poorly motivated that they require some goad or at least some short-term attainable goal in order to work effectively. One-shot end-of-course examinations furnish such motivation.

4. Diagnosis

To identify systems of student failure to learn, comprehend or use knowledge and to diagnose the nature of such failure, is an integral and important part

of a teacher's job. Assessment procedures are the means by which such evidence can be obtained.

5. Evaluation

The recent attention being paid to curriculum development, including the definition of educational objectives, and the search for strategies by which these objectives might be achieved, has led to the application of assessment procedures in the cause of evaluation. The means by which the effectiveness of different courses of instruction on pupils assumed to be similar in their capacities to learn, has been sought by developing assessment procedures which enable curriculum developers to compare the attainments of groups of students treated differently.

It is clear that no one assessment procedure is capable of serving efficiently all these diverse functions. It is necessary to examine the properties of particular assessment procedures and then to decide which procedure best serves the purpose we choose. The two basic properties of assessment procedures have long been identified as reliability and validity.

reliability

The notion of reliability stems directly from the comparison between assessment procedures and any other measuring instrument. In any test, it is theoretically possible to consider a candidate's true score. A true score is rarely obtained in practice because any attempt to measure this true score is subject to error. The distance between a candidate's true score and his observed score is subject to three sources of variability.

The first of these is the variation due to the fact that in any test a selection will have been made both of syllabus content and of the particular skills from a larger universe of content and skills which might legitimately have been used. Thus for example, a Biology Ordinary level paper might include questions on photosynthesis, but no question on respiration; the questions selected may demand the ability to recall information but not the ability to invent hypotheses.

The second source of variation is that due to the unreliability due to the lack of complete objectivity in whatever marking system is used. It is empirically demonstrable that in general terms the longer the answer demanded of a student, the lower is the inter-marker reliability. Conversely, the shorter the answer demanded of a student, progressing ultimately to the case where five answers are given and the student is asked to pick the best of these, the greater the reliability. It is also difficult to establish exact equivalence between two or more tests designed to measure the same attainments. Because of differences of content and of intellectual demand, tests rarely give rise to the same rank order of candidates. Even within a paper, variation between questions is such that the same mark may be achieved by candidates by the exercise of quite different skills, a phenomenon which is exacerbated where choice of questions is allowed.

The final source of variation is that due to variation from time to time in a candidate's performance. Even highly trained athletes, stimulated by the competition on which they thrive, fail to produce identical performances. I

don't know what Davies's true score for the long jump is, but it is clear that on any given day there is no guarantee that this true score is achieved. If this is true for long jump, then it may be even more true for the kind of intellectual gymnastics demanded of examination candidates.

If we examine different assessment procedures in terms of criteria that relate to this concept of reliability, then it is possible to compare their relative merits as reliable instruments for measuring attainment. Thus for example, the argument in favour of multiple-choice objective tests includes the following:

(a) Because it is possible for candidates to answer as many as 40 questions in an hour, examiners can make a more effective job of sampling both syllabus content and a range of intellectual skills.
(b) Because multiple-choice tests demand that the student selects one from four or five given answers to each question, the marking of scripts is a mechanical procedure which may be undertaken by clerks or even machines, which potentially reduces error to nil.

Another advantage claimed for such tests is that they are amenable to fairly sophisticated analytical procedures by means of which items can be selected to make the test a more efficient measuring instrument. The same kind of argument may be rehearsed to accept or reject any assessment procedure.

validity

The second property of an assessment procedure is called validity. I might claim to measure intelligence in a reliable way by measuring the circumference of a person's head. Given that certain simple rules were obeyed, a high degree of reliability could be assured. No matter how often or by whom the measurement was made, the results would show a high degree of consistency. Such a means for measuring intelligence could not however be validated because no connection is known to exist between the circumference of a person's head and what is acknowledged generally to be his intelligence. Validity, then concerns the question, of what we are trying to measure, and whether the instrument we use make demands that are consistent with our description. The first task in the achievement of validity is to define, as accurately as we can, the nature of the intellectual demands that we wish to make on candidates; to define what we are trying to measure. Unlike reliability, validity cannot wholly be established by the collection of empirical data and its statistical analysis. Ultimately, validity depends upon judgement. Perhaps one exception to this general statement is the case of predictive validity. In the case of aptitude tests which are specifically designed for predictive purposes, it is essential to show that present attainment is a good predictor of future performance. Thus an aptitude test is valid if scores in the test correlate highly and positively with some appropriate later measure of attainment. In practice, many public examinations which serve predictive functions have little or no demonstrable claim to this kind of validity.

in-course assessment - advantages

Now let us move on to consider the merits and demerits, both actual and potential, of a species of in-course assessment procedure. In general terms, in-course assessment may be seen to have several advantages over one-shot end-of-course assessments. Of primary importance is the unique advantage possessed by in-course assessment that errors due to variation in a candidate's

performance are minimised. This is so because in this form of assessment information is gathered about a candidate's attainments on a number of occasions throughout an extended time period. A second advantage is that the sampling, both of syllabus content and of the various intellectual demands which might be made on students, may be more comprehensive. Thirdly, the rather special psychological conditions which attend formal end-of-course examinations may be absent.

disadvantages

There are, however disadvantages attending this form of assessment. It may be that teachers find the dual role of teacher and examiner in conflict. In-course assessments conducted by teachers may be subject to what have been called 'halo effects', where attributes of a pupil not relevant to attainment affect the objectivity of a teacher's judgement. Some kinds of in-course assessment seem to depend upon vague, ill-defined, impressionistic and possibly highly subjective judgements and this must militate against both their reliability and validity.

research findings

In a research carried out at Leicester University with the collaboration of 13 schools examined by the East Midlands Regional Examinations Board, a procedure for the in-course assessment of attainment of CSE Physics pupils was subjected to trials. In many ways this exercise proved to be instructive. There is not time to go into all the ramifications of this research, but I will outline its main features and some of its findings. The trial started by attempting to ascertain the particular intellectual skills which teachers consider to be among the more important outcomes of learning Physics at this level. Four of these skills which were regarded by these teachers as of prime importance were selected for further study. These were:

The ability to make inferences from evidence,
The ability to apply facts and principles to problem-solving,
The ability to make observations,
The ability to organise data (in the form of tabulations, graphs, histograms, and the like).

The next step was to discuss with these teachers the means by which a system of intermittent assessment could be devised, in which these abilities could be tested. During a trial experimental period, teachers attempted to identify those episodes in a lesson in which these skills were predominantly used. They attempted to devise the means by which some written and assessable responses might be made by pupils which could be described in terms consistent with one, or more than one, of the four abilities listed above. Thus, for example, a teacher conducting a practical class in which pupils were measuring the resistance offered to the flow of electricity by wires of different lengths and diameters and made of different metals, received results from his class and recorded them in random order on the blackboard. "Test" conditions were now imposed on the class who were invited to try to find ways of tabulating the information, so that certain relationships in the data might become clear. The pupils' answers were committed to paper and the papers were collected. The lesson continued with a class discussion of how best to organise the data. At the end of the lesson, the responses were assessed on a comparative 5-point scale.

In this way attainment data were collected under each of the four headings, approximately ten separate assessments being made under each heading, giving 40 assessments in all, in the class in one year.

Because the system of marking used here was comparative, for example, an average performance, an above average performance, or a well-above average performance, the teacher was able to rank his students one relative to another, but was not able to compare the attainments of his students with those from other schools. For this purpose a moderating device was devised.

The moderating procedure was a test constructed so as to include four sub-tests, each corresponding to the abilities identified by these teachers. Items for these sub-tests were devised both by the teachers involved and by item writers. All the items which potentially contributed to the sub-tests were both vetted by the teachers whose classes were assessed and classified by these teachers into one of five categories. Four of these categories corresponded to the original list of four abilities, the fifth category was a dump for those items which were acceptable but which could not be classified. To be included in the test an item had to be acceptable to at least ten out of thirteen teachers, and must have been placed in the same category by three-quarters of the teachers. The purpose of this test, as with all moderating tests, was to allow pupils placed in rank order by their own teachers to be placed on a common scale. In order to effect this, straightforward statistical standardising procedures were used. It was interesting later to compare the ranks of students as determined by the in-course assessment procedure and their ranks in the moderating test. The correlation between these two was approximately $R = 0.5$. This means that about 25% of the variation in these two sets of scores, the in-course assessment scores and the moderating test scores, could be accounted for by common factors. This may just about be sufficient for the purposes of moderation, but it is interesting to note that 75% of the variation was due to properties uniquely held by the method of assessment used, or to factors wholly outside the control of the investigator. Of particular interest, however, is the relationships which were demonstrated between the students' scores in the four sub-tests and in the four in-course assessments under the four headings used.

TABLE 1

Intellectual skill	organising skill	application	observation
Inference	.24	.34	.17
Organising data		.30	.14
Application			.25

Table 1 shows the correlations between sub-tests. These correlations are low. This tends to suggest that these four skills do not have a great deal in common. It follows that when a pupils' scores on these four sub-tests were added together to give some global score for attainment in Physics, there was, to say the least, a considerable loss of information. It is highly likely that whenever global test scores are used to describe attainment, there is a similar loss of information. It is this kind of evidence which points to the conclusion that whatever system of

assessment we use, profiles will furnish a better description than either scores or grades.

It is interesting to note that the teachers who were involved in this exercise, although at first somewhat sceptical, later admitted that for them the most important side-effect of this trial was to make them more aware than they had previously been of their students' strengths and weaknesses, and of the success or failure of some of their teaching. The purpose of this paper is not, however, to persuade you that in-course assessment conducted along these lines provides all the answers to problems of reliability, validity and of ameliorating the adverse effects of one-shot examinations. It is that in-course assessments, whatever further enquiry needs to be done, have a potentially important part to play in the assessment of attainment and that they may furnish a useful weapon in an armoury of assessment techniques."

Dr G D Moss,
Director,
Centre for
Educational
Technology,
University College,
Cardiff

EXAMINATIONS AND ASSESSMENT AT THE OPEN UNIVERSITY

"In any system of assessing and examining students the assessor must have clearly defined aims or a set of aims in mind. It is in fact possible to identify a series of aims which can be applied to assessment systems where no single system is likely to fulfil every single aim which can be specified. A list of possible aims of assessment are given in Table 1.

TABLE 1

1. To grade students and rank them according to their level of achievement compared with one another, ie a Norm-Based test.
2. To measure the student's mastery of a subject in relation to a set of clearly defined teaching objectives, ie A Criterion-Based distribution.
3. To analyse the pattern of student responses in order to diagnose good and bad test items.
4. To use the pattern of student response to diagnose areas of student weakness or of ineffectual teaching.
5. To use patterns of responses to identify inherently difficult areas of subject matter within a course.
6. To return marks or grades to students in order to give them a measure of their own progress, ie an incentive.
7. To return details of marks together with answer schemes to students in order to permit the student to diagnose his own areas of weakness and take appropriate remedial action. (Many people now feel that if such a diagnosis could be carried out by a computer then a remedial process might be provided by the computer direct to the student).
8. To add together the marks from a variety of tests in order to award some form of overall credit.

The aims outlined in Table 1 cannot all be achieved by the same assessment model. There are in fact, two basic models in use which probably represent opposite ends of a continuous spectrum. At one end we have a system of examining in which the objective is to grade students in a Norm-Based manner. Characteristically, the tests are Norm-Based, no feedback is given to students and scores may be adjusted after marking. Such tests achieve aims 1, 3, 5, and 6, and also enable part of aim 4 to be achieved - that of diagnosing areas of student weakness. At the other end of the spectrum are tests containing Criterion-Based or mastery items in which a high correct response rate is hoped for and much feedback is given to students. There is no point in making any adjustment to the scores gained on such a test and the aims which are covered are 2, 3, 4, 5, 6, and 7. Notice that at this point I have deliberately excluded mentioning aim 8, although I shall return to this at a later point.

the open university system

The Open University system of examination and assessment varies from course to course and from Faculty to Faculty but there is a basic element which is similar on all courses. An Ordinary Degree can be obtained by qualifying for six full credits, two of which must come from foundation courses and four from post-foundation-courses. The length of a normal one-credit course is thirty-four units or in terms of hourly equivalents, approximately 400 hours of work. The commonest form of course which is offered in The Open University is a half-credit course and I shall use the half-credit course as the model for further analysis of the assessment and examination system.

The half-credit which is awarded on a half-credit course is composed of a continuous assessment element and an examination element and this is usually awarded on a 50/50 basis. The 50 per cent which comes from continuous assessment comes from a combination of computer marked assignments (CMAs) and tutor marked assignments (TMAs). The relative proportions of CMAs and TMAs varies from course to course and from Faculty to Faculty. In the area of Science and Technology, the normal allocation is shown in Table 2.

TABLE 2 Continuous Assessment

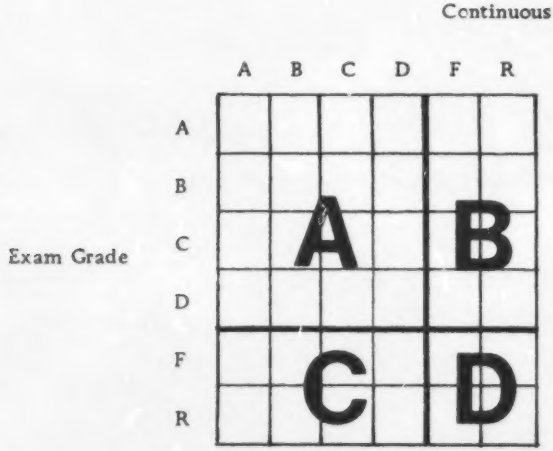
8 Computer Marked Assignments (see App. I)	25%
4 Tutor Marked Assignments (see App. II)	25%
Examination	
1 Computer Marked Section	25%
1 Tutor Marked Section	25%
TOTAL	100%

In terms of work load the continuous assessment portion of the assessment takes students approximately 32 hours out of the 200 hours devoted to the half-credit course. Each CMA takes approximately two hours while each TMA takes approximately four hours. The examination portion is a single three-hour examination.

The assessment is carried out on course objectives and unit objectives so that the three-hour examination is based on the course objectives, while the CMAs and TMAs are based on the unit objectives. Care is taken to ensure that the questions which are asked both in the continuous assessment and in the examination cover as many of the objectives as possible and cover them at as many levels as are possible. Thus we try to avoid constructing test items which are based purely on factual recall but attempt to produce items which depend upon the higher learning skills such as analysis and interpretation of data, the interrelationships and applications of high level concepts, and the application of innovatory skills such as the design of experiment in the area of Science and Technology.

The award of a credit is based not upon the simple addition of the overall continuous assessment grade and the overall examination grade, but instead the students are ranked in a matrix such as that shown in Figure 1.

FIGURE 1 The Open University Conflation Matrix



In Figure 1 all students in box A, have clearly passed both the continuous assessment and the examination component and are automatically awarded a credit. All those in box D, have clearly failed both components and are not considered for a credit unless extenuating circumstances have prevailed during the period of study in that course. Those students who fall into the areas (B) and (C) have passed one component of the system and failed in the other. These are designated as borderline cases, and are arbitrated upon by the Examination Board which of course, includes external examiners. In many courses in which there is a Summer School component then the Summer School grade is used to arbitrate between students who fall into these borderline categories.

In practice we are only just learning how to interpret the results of an overall continuous assessment grade. Such a grade is arrived at by averaging the grades obtained by the students on their CMAs and TMAs; obviously any student who obtains an overall (A) grade is an exceptionally good student indeed and an overall grade of (B) should be regarded as highly acceptable.

operational problems

Certain operational problems exist:

1. The specification of objectives in higher education is extremely difficult and time-consuming and it's probably true to say that the absolute specification of such objectives would result in pages and pages of objectives for every course unit written. As a result, each major objective is specified in only one or two areas relevant to the unit. Consequently, the adoption of any kind of Criterion-Based test is relatively difficult based on such an objective. However, the specification of certain low level objectives is much easier and the resulting situation is one in which it is relatively easy to produce Criterion-Based tests on low level or basic objectives while the attainment of the higher level objectives is measured using Norm-Based tests.

The problem is that Norm and Criterion referenced items are frequently inadvertently mixed within CMAs and TMAs. Now the kind of grade distribution which is given by a Norm-Based test and by a Criterion-Based test are completely different, and it is quite invalid to herd together the scores from the different kinds of items.

2. The whole organisation is complicated by a consideration of political issues. If we could specify our objectives and write valid criterion referenced tests for all the objectives and if we could obtain a 95 per cent pass-rate on such tests (which is not atypical for a criterion referenced test) then such a result would probably be as unacceptable, politically, as a 95 per cent fail-rate. We are therefore forced to accept a situation which is less than perfect in this respect.

3. Cheating or co-operation on the continuous assessment system does go on. Students who do their assignments under unsupervised conditions in their own homes are known to co-operate in working out the answers. In a system such as the Open University it is impossible to eliminate this aspect of the continuous assessment system, however, it can be pointed out that those students who do well on the continuous assessment and enter the examination room with a pass-grade on continuous assessment almost invariably pass the examination which is carried out under strict conditions employing normal examination procedures.

4. One of the biggest problems in the system at the moment is that in any given half-credit course the student is allowed to submit a minimum of six CMAs out of eight and three TMAs out of four. His continuous assessment grade is calculated on his best six CMAs together with his best three TMAs. The problem is that not all of the CMAs and TMAs are of equal difficulty, ie they are not equivalent and the element of choice which is built into the system means that different students complete different combinations of assignments and yet these marks are added together to form an overall continuous assessment grade which is used to compare one student with another.

5. Because the Open University is a young and emergent organisation, there is an absence of pre-tested assessment items. This is currently being corrected by the implementation of a computerized item analysis system. It is desirable to test the items before they are used in course assessment but this has proved difficult in terms of finding suitable populations of students and large enough populations of students on which to carry out the pre-testing.

6. The tutor marked section of the assessment system carries with it an almost guaranteed unreliability which is the variation between markers. The system operates so that each tutor markes approximately twenty scripts from students on any given course. Obviously on a course with several thousand students there is difficulty in getting tutors to mark to a standard pattern, even when standard marking schemes are issued.

This problem is largely avoided during the script marked examinations by using much smaller panels of markers marking larger numbers

of scripts. In such a situation we know that the normal standardization procedures characteristic of most examination boards are carried out.

advantages of the system

The system has three advantages:

1. A wide variety of computer scoring options permit a wide variety of test formats to be employed in the computer marked assignments. Such formats range from the simple multiple-choice items with a one-from-several design to the very complex conceptual group coherence problems which are capable of testing the highest learning skills.
2. The overall computer control allows analyses of response patterns which leads to question and course revision and it also ensures that the marking is standard.
3. Computer control permits automatic adjustment to be made to script marker scores or to the distribution of scores in other exams and tests. The adjustments to the script marker's scores are made after monitoring procedures have been carried out.

future possibilities

I have pointed out the problems involved in adding together marks from tests with varying degrees of difficulty and the matrix system which is used to get over the problem of adding together marks from continuous assessment (based on unit objectives), and examinations (based on course objectives). However, the problem is deeper than that in that it is one of measuring the attainments of certain skills which are specified in the course and unit objectives. By adding together the marks which represent the performance on a variety of items within a test, and by adding together the marks from a variety of tests, we are in fact combining all of these marks to give one overall grade of doubtful validity and a grade which gives little information as to which skills and abilities the student really has. The system we are hoping to move to in the future is one in which Criterion-Based elements will be clearly distinguished from non-criterion or Norm-Based elements and the assessment carried out in separate tests. Further than this we hope to be able to clearly distinguish student performance in different skills and to avoid adding together marks which reflect abilities in different skills. The implication of this is that for each student on each course we would produce not a single overall grade, or a point on a matrix, but a student profile based on the skills taught in that course. Such a hypothetical situation is shown in Table 3.

The advantages of such a system are that the scores in the profile will be useful not only to the student, but to a potential employer, or to someone who is assessing whether the student is capable of going forward to do a higher level course.

The student can clearly see those skills and abilities which he has succeeded in attaining, and those skills and abilities in which he is weak. If he really wishes to develop the skills which are weak, then he can direct his attentions to remedial study in these areas. Similarly, potential employer can assess the student's abilities as measured by this more revealing system and can decide accordingly as to the suitability of the candidate. The overall point

to be made is that in any kind of assessment system in which large sums of money and a great amount of time and effort are devoted to the measurement of certain skills and abilities, it is ludicrous to then add together marks in a meaningless way so that valuable information is literally thrown away for the sake of obtaining a single overall number or letter grade of doubtful validity."

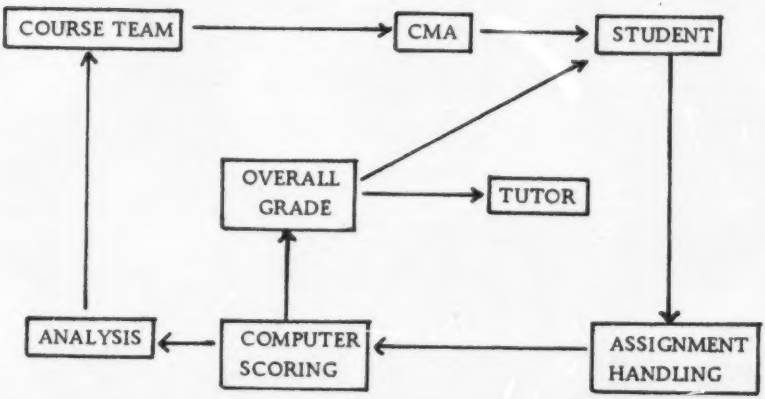
TABLE 3 A profile assessment based on acquired skills related to specific objectives in a science course.*

Ability/Skill	Test 1	Test 2	Test 3	Test 4	Overall Profile
Factual recall	A	A	A	A	A
Measurement and estimation of error	B	N/A	C	D	C
Recording of results	N/A	B	B	N/A	B
Analysis of results	F	B	N/A	D	F
Interpretation of data	C	D	D	N/A	D
Design and innovation	N/A	N/A	B	C	B/C

* Grades awarded on A - R scale. N/A indicates topic not assessed in that test.

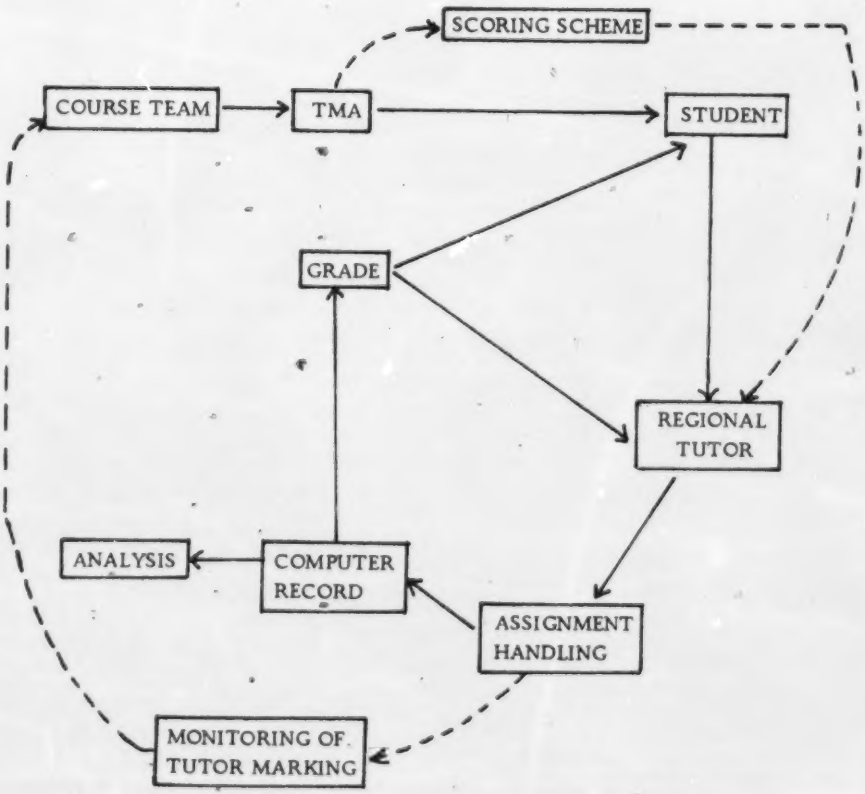
APPENDIX I

CMA



APPENDIX II

TMA



Mrs C Ward,
Examination
Techniques
Developments Unit,
CGLI

DEVELOPMENTS IN CGLI EXAMINATION AT CRAFT AND OPERATIVE LEVELS

OBJECTIVE TESTS

It would be impossible to give this paper without speaking of objective tests, the most familiar type being the multi-choice question with a key and three distractors.

advantages

There are seven main advantages of objective testing:

(a) Good Syllabus Coverage As there are usually 50 to 80 items in an objective test the syllabus can be covered in a systematic manner. A specific number of questions can be allocated to each section of the syllabus. We try to test three categories of knowledge in each section: factual recall, comprehension and application of knowledge. When devising examination papers, we try to avoid testing only factual recall although, of course, these are the items which are easiest to write. With the help of our item writers we have been able to increase our testing of comprehension and application of knowledge.

(b) Uniform Syllabus Coverage If all questions are compulsory, which is possible in a multi-choice test, all the candidates take the same examination. In a traditional examination paper where candidates answer, for instance, six out of 10 questions, 210 different combinations are possible. It is difficult to devise a traditional examination in which all the questions are of equal difficulty and research has shown that candidates do not always choose to do the questions which they will answer best.

However, it is necessary in certain circumstances to provide an element of choice in objective tests, for example where there are certain options in the syllabus. In such cases the test is divided into sections and the candidate must answer all questions in a specified number of sections.

(c) Reduced Emphasis on Penmanship Examinations, except in English, are not intended to measure a candidate's ability to write. Objective tests enable a candidate to demonstrate his knowledge unhampered by, perhaps, a low degree of proficiency in writing. Objective tests, cannot, of course, avoid testing a candidate's reading ability as well as his subject knowledge.

(d) Reliability Marking In a well constructed objective question there is only one right answer. If any others inadvertently appear they are eliminated

in the pre-test stage.

(e) Statistical Analysis In this respect objective tests have two advantages. Firstly, questions can be pre-tested so that we can see whether they do in fact work. We can analyse their facility, discrimination and the effectiveness of distractors, although we try to get this right before the pre-test stage. The pre-test enables us to identify and eliminate questions which are unsatisfactory so that the examination contains only satisfactory questions.

Secondly, the standard of the test can be kept at the same level. Therefore we can compare performance from year to year. Our experience is that the analysis we get for a question on pre-test is comparable with that of the examination itself. We can then say with confidence if there is a change in the results, that it is due to a variation in the students, and not in the examination.

(f) Rapid issue of results is possible where the examination contains only objective questions.

(g) Increased Opportunity for Teacher Participation More teachers and industrialists are able to participate in the examination process, eg in item writing, item editing and item review. We distribute comment forms to colleges using our multiple choice pre-tests and invite comments on the examination as a whole or individual items. These can then be taken into consideration when the items are reviewed and help in deciding whether they are satisfactory. Similar comment forms are used for multiple choice examinations.

developments

There have been certain developments and refinements in objective tests since the CGLI began to use them:

(a) Objective tests need not consist only of multiple choice questions. Alternatives are multiple response questions, matching block questions, assertion/reason questions and true/false questions. We avoid using true/false questions at CGLI because a candidate has a 50% chance of being right simply by guessing. We have experimented with the other types but converted them to multiple choice format so that they can be marked in the same way as other items.

The various types of item may be illustrated by examples taken from the Alcoholic Beverages Service Certificate. In a multiple response question the candidate may be asked to indicate which two of several possible factors combine correctly to define a particular phrase. In matching block questions the candidate has for example to match a list of types of wine with names of wine. In the assertion/reason type of item a statement is made, eg 'a wine waiter would normally recommend sauterne with clear soup because it is a sweet white wine'. The candidate is asked to say which of the following is true

(a) both assertion and reason are correct and the reason is a correct explanation of the assertion
(b) both are correct but the reason is not the correct one for that assertion

(c) the assertion is untrue
(d) the reason is untrue

These types of item give more scope for asking different types of question but there is a danger of the candidate being confused by the actual form of the question and being tested on his comprehension of this, rather on his knowledge. Therefore the CGLI approaches this type of question with caution and uses them only when there seems to be no other method of getting the information from the candidate.

(b) For some time use has been made of group questions, a set of questions based on a single diagram or piece of information or statement. The advantage of this approach is that the candidate's attention is concentrated on one particular problem. For instance, in a Radio and Television Service Examination questions of varying difficulty may be related to a diagram of a circuit.

(c) It is possible to devise permutation sets for examinations using objective testing. Before writers actually compose the questions they make a list of what the candidate is expected to know. From this questions will be drawn up and possible permutations of distractors selected. This method does not work for all syllabuses but is useful in certain subjects.

uses Objective tests may be used for:

(a) End of term tests.

(b) In-course testing This method of using objective tests originated with the Motor Vehicle Part I course in Scotland. Tests are given throughout the course as students complete each block of work rather than in a single final examination at the end of the year. Advantages to the teacher are that there is continual feedback, no necessity for a fixed examination date and the possibility of relating some of the blocks to different material to suit individual needs. One suggestion made, for example, is that on the Basic Engineering course craft practice blocks might be substituted for students who are not receiving on the job training.

This usage does have disadvantages, however. Administrative problems are considerable even where the bank of items is sufficient. There are difficulties in getting the tests assembled and distributed to colleges at the right time and so on. Another problem is that the items are taught in separate compartments and so the students do not necessarily relate them to each other. The system is also criticised for not requiring students to retain their knowledge of each block they study throughout the year. Some courses therefore have a short final test of 20 items which attempts to link and relate the various component blocks. It contains, in addition to revision items, items with distractors drawn from different blocks. Experiments with the system are being tried in a few areas and if they prove successful the system might be extended.

(c) Entry Tests CGLI is experimenting with giving selected questions from the end-of-year test to students at the beginning of their course. In the experiment

carried out with Motor Vehicles Part I students in three colleges some did very well - 45% would have passed the final examination and one candidate would have got as much as 92%, indicating that it would be a waste of his time to take the course at all and that other candidates would not reap full benefit from a year's study. It was found that some candidates got full marks on particular sections but did less well in others. The CGLI wonders whether it would therefore be a useful service to colleges to provide an entry test, especially for those with students coming from link courses, in order to determine those areas where a student will not need much tuition.

Secondly, we intend to compare students' performance at the end of the year with their results in the entry test to see if we can gain any clues as to how a candidate is likely to do in the course judging from the performance in the initial test. We shall look in particular at the significance of their performance in science and calculation items in the initial test. In our sample there were four such items in the entry test which the candidates did better by 1-2% than did candidates taking these items at the end of the year. It should be stressed, however, that the use of objective questions for tests other than end-of-course examinations is still in the experimental stages.

TRADITIONAL TESTS

Having surveyed the reasons for using objective questioning and developments which have taken place in this type of test, perhaps we could now consider the traditional type of test.

CGLI annually sets about 550 different craft and operative written questions and a considerable number of these are not objective items. Some of the traditional tests are in declining schemes, eg craft practice, but for some there seems little likelihood of their being replaced by objective testing, either because of the subject matter or because there are a comparatively small number of candidates and it is not economic to devise objective questions.

CGLI is studying the performance of traditional type examinations, trying to improve them and to draw on our experience of objective testing and apply some of the disciplines of objective testing to traditional examination papers. Recently we have published a booklet which gives guidance to our examiners and moderators in setting traditional papers.

improving performance

We can consider improvements in traditional examinations by considering the aspects listed earlier as advantages of objective tests.

(a) Syllabus Coverage It is not easy to achieve good syllabus coverage with a traditional style examination paper consisting of 10 essay type questions of which the candidates have to answer six. The amount of syllabus content which can be covered in such a question is restricted. If, however, shorter answers are demanded a somewhat wider range of the syllabus can be tested in the examination. Another possibility is the use of structured questions, rather similar to grouped questions in objective tests, in which a number of short questions are related to a central problem or diagram.

We are considering the possibility of establishing a specification for traditional examination papers, deciding how much weight should be given in the examination to each syllabus section. We have also tried to allocate different parts of the examination to different types of knowledge, as we do in objective testing, but it has proved very difficult to establish what types of knowledge are tested in each question. Perhaps we should look at the type of answer required viz. diagram, essay. Work on specifications for traditional papers is still only just beginning.

(b) Uniform Syllabus Coverage and Elimination of Choice It is unfair in a traditional examination demanding a few essay type answers to expect the candidate to answer every question on the paper. If he is unable to do one out of, say, six, he loses a sixth of the possible marks. But if short structured questions are used, such a demand is more reasonable and it is then possible to ensure that all candidates are answering the same set of questions. Again, account must be taken of the possibility of options and specialisations in the syllabus and the length of the syllabus. Sometimes it is too long to be taught in the time available and colleges do not cover every topic.

(c) Emphasis on Penmanship The need for penmanship cannot be eliminated but there is an increasing tendency to test descriptive matter by means of objective questions, reserving traditional question papers for answers which demand calculations or the drawing of a diagram. Even when candidates have to write an answer, short or structured answers reduce the need for an essay-type response. This means that candidates are not penalised if they cannot write quickly or express themselves well in continuous prose.

(d) Reliability of Marking Traditional examinations cannot match the reliability of objective tests but if the short or structured answer is demanded, it is easier to mark and can be marked more reliably than the essay-type answer. In a traditional essay-type answer the candidate may not know exactly what he is expected to include in his answer and the examiner himself may not even know. If, for instance, he decided to give full marks to a candidate who makes three points what does he give the candidate who makes five? There is a considerable degree of subjectivity in marking. If questions are worded specifically, reliability is easier to achieve. We are trying to ensure that examiners produce detailed marking schemes and submit them for moderation with the draft examination papers. If precise marking schemes can be produced, reliability improves. CGLI tries to get examiners to co-ordinate with each other, discuss their marking schemes and consider the results when they all mark the same script, to see whether they are all marking to the same standard. Experiments have been carried out to try to identify what type of question produces the most reliable marking.

(e) Statistical Analysis The ability to pre-test questions, establish their level of difficulty and compare results from year to year is one of the main advantages of objective testing. Research is being carried out on the analysis of traditional questions to see how far statistical analysis can be applied to them. The Educational Measurement Research Unit at Reading University has pioneered this work and the CGLI has worked with them on CGLI questions.

Aspects studied are the position of a question on the paper (candidates may have insufficient time for the last question), choice index (percentage opting to do that question), the mean mark on the question, the standard deviation (indicative of the spread of marks on the question), "mean total" (indicative of how well the candidates who answered the question did on the whole paper), facility index and discrimination index. The facility index involves a comparison of the mean mark obtained on the questions with the ability of the candidates answering the question (shown by the "mean total" figure) and assuming that an average candidate would obtain 50% on an average question, marked by an average examiner. Together with the mean mark and discrimination index this indicates how successful a particular question has been. The analysis thus obtained is discussed with the moderating committee for the subject with a view to improving future papers.

Therefore opportunities exist for further work in both objective testing and traditional examining and we hope that by applying the disciplines of objective testing to traditional examinations we may effect further improvements in this area. "

A Taylor,
Head of Computer Services, City and Guilds of London Institute.

DATA PROCESSING OF CITY AND GUILDS EXAMINATIONS

"Further education is in a process of continuing change to meet the changing needs of industry. Mrs Ward has said something of what the Institute is doing in the development of assessment methods in this changing situation. We will now look at the data processing which has been designed to assist in these assessment methods.

computer systems

The Institute uses four computer systems to:

(a) process examination entries
(b) allocate scripts to examiners
(c) mark objective (multiple-choice) papers
(d) process examination results

Our discussion will be concerned with the last three of these systems. The first of the systems has been mentioned for completeness since for the use of any assessment method there must be candidates entered for assessment.

script allocation

Even with the introduction of objective papers, the majority of candidates' written work is still marked by examiners - some 440,000 scripts were submitted in the May-June 1973 examinations. Procedures for the marking of scripts are clearly vital to success in the assessment methods; one of these is the allocation of scripts to examiners.

From the computer processing of examination entries the system for the allocation of scripts prepares and prints lists showing for each examiner the colleges from which scripts are to be marked and also addressed labels for sending the scripts. This means that as scripts are received from colleges they can be sent without delay to examiners, thus keeping to a minimum any delay in starting marking. A complete record of scripts received and despatched is thus maintained.

marking

For written papers the marking is done by examiners appointed by the Institute; objective papers are marked by computer. Practical tests are marked by examiners appointed by the Institute. Course work and other college assessments (eg written papers set by the college) are marked by persons designated by the college subject to assessment as required by the Institute. Marking schemes and instructions are issued by the Institute or in the case of college set and marked assessments are agreed with the Institute. The answer sheets completed for objective papers are 'read' by an optical

scanner and 'marked' by comparing the answer 'read' with the correct answer. Marks for all papers and assessments are input to the computer system for processing results.

analyses

Analyses are prepared by computer of the various marks. These analyses include:-

(a) mark analysis and distribution,
(b) result analysis,
(c) item analysis and test statistics (objective papers only).

They are used by the Institute, in consultation with examiners and awarding committees as appropriate, to determine the final results to be issued.

checks

The computer systems have various checks inbuilt, eg to ensure that the marks are credited to the correct candidate and for the correct paper. As, if not more, important are the checks to ensure that the marks are correct; these include the checking of scripts marked by examiners and of the computer print-out from the marking of answer sheets (for objective papers).

results and statistics

After final consideration of the calculated results, computer printing of the results for issue to candidates and colleges takes place. This is followed by printing of certificates for successful candidates and the final statistics for the Institute's records.

B Argent,
Administrative
Officer,
CGLI

IDENTIFYING FACTORS FOR ASSESSMENT

"My brief is to speak on 'identifying factors for assessment'. I have taken this to mean ways of locating and selecting what should be assessed, and becoming aware of other factors such as time constraints and priorities which may exert considerable influence on the choice of what we actually do assess.

Before moving onto the business of identifying factors for assessment, I would like to lay my cards on the table regarding two central issues, namely:

1. what I understand by the term 'assessment' and
2. how I see the relationship between assessment and teaching

what is assessment?

On the meaning of the term assessment I do not agree with those who accept or insist that assessment of student achievement is something different from setting and marking tests or examinations. This approach usually implies that there is something specific, perhaps limited, and accurate about a test, whereas assessment is less precise. Such reasoning creates an entirely false and unproductive division, and it seems to me that a more effective way of looking at assessment of student achievement is like this:

Assessment of achievement

..... means (1) of judging (2)

student success or failure (3)

This is not a definition of assessment, but it contains to my mind features which we shall find useful to keep before us.

Firstly, assessment of achievement may be accomplished by any single method or combination of methods. The choice will depend on what achievements are to be assessed, and other factors such as how much time is available for assessment, and how skilful and knowledgeable in assessment techniques is the person making the assessment.

I stress this point because the work necessary to decide what it is you wish to assess may well be wasted if you are unable to select and construct the appropriate assessment methods. By assessment methods I simply mean all those means which

call upon students to do something and involve someone judging what they have done, how they do it or both.

For example a test of 30 true/false questions is an assessment method, and so is a record of punctuality and a rating scale of attentiveness and participation in class.

Secondly, it reminds us that judgment is an essential aspect of assessment.

Judgment comes in three times in any assessment method:

(i) at first when we decide what it is we wish to assess,
(ii) then when we construct assessment methods whether it is writing multiple-choice questions or devising an attitude questionnaire, we exercise judgment in our choice of words, situations, etc.
(iii) finally any application of an assessment method concludes with a judgment about the students' success or otherwise in reaching the required achievement.

Returning to our outline of the basics of assessment, we see that it reminds us these final judgments are always essentially of the same kind; what are sometimes referred to as 'Go'/'No Go' decisions. That is however we 'dress-up' or camoflage our judgments as marks out of 10 or 100, or as grades such as A, B, C, or α, β, γ, or other 'mystical' symbols, we must not forget that our fundamental choice is between deciding 'Yes - that's good enough' or 'No - that's not good enough'.

assessment

I hope that I have managed at least to give a coherent even though rather truncated view of how I see assessment. Because this view leads me to the relationship between assessment and teaching, which is that assessment is a natural and essential part of teaching, if by teaching we understand the process of directing and assisting the learning of others. Fortunately in further education most of our courses have, or can be provided with reasonably clear cut targets for learning and I doubt if many would disagree that finding out about students' progress towards those targets is a necessary part of a teacher's job, whether it be during learning when he is concerned with determining strengths or weaknesses or at such times when he has to decide whether the student has 'passed' or 'failed'.

success or failure

My 'basics of assessment' ends with students' success or failure, and this brings us to our main theme success or failure. In what?, or rather how can we go about identifying 'the what'.

factors to be identified

Well I believe such identification can be engaged upon effectively if we recognise that these are several different categories of factors to be identified, factors associated with:

1. course content,
2. external demands,
3. priorities,
4. individual needs,
5. assessment capabilities.

These five categories are not independent of each other.

Let's begin by considering identification of course content factors. On these the NEBSS pamphlet is very direct, but perhaps not very forthcoming. On page 8, we are told that the purposes of the NEBSS scheme are.

1 ..... to impart a sound knowledge of the principles of supervision
2 ..... to help potential supervisors and supervisors get the feel of applying these principles in practice.

These statements identify two major classes of factors to be assessed, ie successful achievement of:

1. a sound knowledge of the principles of supervision - CLASS I
2. the feel of applying these principles in practice - CLASS II

The accusation may be levelled at me that I have not identified these two classes at all but merely discovered them. I accept that, and we may consider by whom and how they were identified.

My line of enquiry has taken me to the following tentative conclusions.

Clearly they were identified by a group of people cognisant with the needs of supervisors, and the part education and training might take in fulfilling those needs.

I expect the 'how' was largely a matter of adherence to common sense, experience and a bit of educational philosophy. It is accepted educational thinking that many activities can be understood and practised provided one acquires a sound knowledge of the principles of that activity and benefits from guided experience in applying the afore-mentioned principles as appropriate.

Now I do not know that this was the process by which our two classes of factors were arrived at, but sometime or other I believe we need to know on what basis they were obtained because it is achievement of these two, which needs to be assessed. And equally important, we should avoid assessing anything else, or at least avoid allowing our assessment of other factors to affect our judgment of whether or not a student has reached the required standard in:

1. his knowledge of the principles of supervision,
2. his feel for applying these principles in practice.

I should like to stress this because not assessing irrelevant factors is just as important to valid assessment as is assessing relevant factors. Perhaps it may seem that knowledge of the principles of supervision is a bit on the broad side, and that assessing someone's feel for applying principles is also a little vague. To me they represent pretty difficult tasks, which verge on the impossible if nationally comparable standards are to be maintained between colleges.

So our next move is to look at ways of identifying in more detail the constituent factors in:

'a sound knowledge of principles'
and
'a feel for applying principles in practice'.

Once again we can begin by consulting the NEBSS pamphlet, and on page 24 we find FIVE areas of knowledge which are recommended as essential foundations for a course:

1. Principles and Practice of Supervision,
2. Industrial Relations,
3. Technical Aspects,
4. Economic and Financial Aspects,
5. Communication.

Now presumably within each of these areas there are principles relevant to supervision, and we could identify possible assessment factors by itemising such principles in each area for example in 'communication' principles relevant to the supervisor are:

1. Human attention is selective,
2. 'Expectancy' influences reception, and so on.

We could then go a step further and for each principle we could note the type of performance, or manner or behaviour we expect of the supervisor who is applying it in practice for example, for the above two principles we can list:

1. Prepare instructions and briefings beforehand,
2. Give instructions and briefings in an appropriate language and manner,
3. Ask questions to check on understanding and reception, and so on.

When both steps are complete we have much of the relevant information to decide

1. what factors in knowledge of principles should we assess
and
2. what factors regarding a student's success or otherwise in applying these principles in practice should we assess.

This two step approach can be applied to whatever content is chosen as appropriate for the course.

Now I must confess some misgivings about the theme I have adopted so far, because I have quite deliberately treated indentification of factors for assessment from inside the accepted ideas of course aims and content, and have not concerned myself with that aspect of identification which is to do with whether the content of the course really is suitable for the course members.

external demands

In several places the NEBSS pamphlet refers to job analysis and provision in course of different emphases to meet the requirements of supervisors from

different industries and firms, and presumably colleges running NEBSS courses do have recourse to job analysis or can at least make use of job descriptions and job specifications which firms already have.

Presumably such specific external demands will lead to the identification of factors for assessment which differ between students or rather they provide the basis for different emphases on the same factors. Since in a typical class each member may require his own particular mix these external demands provide identification of the factors and relative weighting of factors in the project report, course work, and associated oral assessment.

frequency and criticality

Similarly there are some factors associated with priorities; time limitations also mean that assessment must be selective, which in turn means that we should concentrate our assessing on those factors which must be assessed. Our best guides to this are estimates of frequency and criticality.

For example for each of our lists of principles we could obtain ratings of how frequently the supervisor's work requires him to have knowledge of a principle and how critical his correct application of that principle in practice is likely to be.

Principles of Supervision		
Description	Frequency	Criticality
1.		
2.		
etc.		

Illustrative frequency and criticality ratings

Frequency	Criticality
5 - every week	5 - likely to cause an accident or walk out
4 - once a month	4 - likely to hinder production
3 - once every 3 months	3 - likely to make supervisor unacceptable
2 - once every 6 months	2 - likely to make supervisor ineffective
1 - once a year, or less	1 - likely to 'irritate' some employees

I mentioned that these categories of assessment were inter-related and this fact is especially true of No. 4 'individual needs'.

general assessment factors

In keeping with the general NEBSS philosophy we can see how the planning of course content and subsequent identification of what is to be assessed requires reference back at all points to the course members industrial and company base, there is in addition the matter of his personal base, that is his strong and weak points of personality, previous educational attainments, etc. Not in a general sense, but as they relate to his chances of success on the course. As far as identifying general assessment factors goes we might consider such things as - basic skills in reading, writing and discussion, study habits, use of reference material, etc.

NEBSS assessment

Now let us turn to the last of our five factors and the one that perhaps stands most on its own, that is the assessment capabilities of those responsible for NEBSS courses. I believe we have here a most relevant crop of factors because in spite of identifying all sorts of relevant and critical things to assess we in all probability end up assessing only those which we have time for and only those which can be assessed by methods with which we are familiar.

Most explanations of human learning lay great stress on knowledge of results, and assessment is the teacher's tool for providing that knowledge, and that developing proficiency in assessment techniques is only recently being seen as a desirable feature of teacher training courses. What I can suggest is that taking stock of both the likely efficiency of different assessment methods, and personal skill in using the methods is a good place to start. A checklist like this might be used:

Achievements to be assessed	Methods to be used	Familiarity with confidence in method

May I end with two points. One an admission of incompleteness and the other an overall viewpoint which I hope you will share.

expressing factors for assessment

Regarding incompleteness I must admit that I have said very little about how to express factors for assessment, and it can be argued that the form of expression is extremely important.

Let me illustrate this by reference to one of our two main factor groups

'sound knowledge of the principles of supervision'
what exactly does sound knowledge mean -

Does it mean ability to distinguish between correct and incorrect versions of the principles to given situations or ability to predict the likely results of failure to follow certain principles?

The concept of learning objectives, that is statements in terms of specific, observable behaviour of what a student should be able to do if his learning has been successful is familiar.

Here is an example based on one in a useful little book on the subject 'Stating

behavioural objectives for classroom instruction' by N E Gronlund.

Achievement: 'understands supervisory principles'

The student should be able to:

1. State principles in his own words
2. Give instances of each principles in action
3. Select which principles apply in given situations
4. Distinguish between sound and unsound applications of principles, and so on

I would suggest that being able to give clear statements of learning objectives like this is an essential aspect of teaching as well as a prerequisite for assessing.

assessing personal characteristics

The examples I have given so far have been restricted to what the theorists call cognitive skills, eg remembering, understanding, applying, analysing, etc., but I believe NEBSS courses aim to develop other qualities as well, qualities which the theorists call affective, but which in every day life we look upon simply as personal attributes, such as attitudes, manner, etc.

For example 'dependability' is looked upon as a good quality, as is willingness to consider new approaches to old problems, whereas an erratic or firey temperament is probably considered unwelcome in a supervisor.

Identifying such personal characteristics for assessment is probably just as important as identifying the mental factors, but a cautionary note is necessary. In other fields, eg management development, teacher training, salesman training, it has been found extremely difficult to identify good and bad personal qualities, largely I believe because people are a mixture of these qualities, and a certain mix may work for one person, and a different mix for another. Caution is also required in judging these characteristics, obtaining valid and reliable measures of them is usually very difficult.

My overall viewpoint is this. Identifying factors for assessment is not something we do just to be able to defend our examination papers, tests, rating schemes and so on. We do it because it is part and parcel of teaching, an essential if we want learning to be effective and efficient. In a system like NEBSS it need not, and should not be engaged upon by individual teachers or individual colleges. This would be a duplication and waste of effort. Rather we should make the best of our resources and time by pooling them on regional and where appropriate on national bases. I suspect we shall find that there are many common factors to be assessed and if so we could work together toward producing and using common assessment methods."

The Further Education Staff College was established on 15th November, 1960, with the principal objective of improving the efficiency of further education establishments. Since 1963 the college has been organising study conferences at which senior personnel in education, industry and government are given the opportunity of discussing and analysing developments in further education, and of making recommendations for action and enquiry.

Study conference members meet both in plenary sessions to hear and discuss lectures and in syndicates to discuss specific topics and projects. Syndicate discussion may be augmented by the employment of expert witnesses, visits to other establishments, and simulation exercises in administration and personnel management.

In 1963 the college undertook the publication of Session Papers outlining the work of conferences, and in 1967 the College Report Service was instituted to publish documents on a number of issues relating to tertiary education.

Within its Report Scheme, the Service publishes twenty Reports each year (forming one volume), which include accounts of the proceedings of selected study conferences, collections of papers and special reports. Subject matter covers the development of the objectives and instruments of further education; the inter-relationships between colleges, industry and other branches of education; specialised subject developments; college administration and management.

These Reports do not necessarily express the views either of individual conference members or of the Staff College.

From Volume 4 onwards, a subscription may be made to the Scheme at £12. per annum. Further copies of individual Reports may be purchased by subscribers at 60p each, and non-subscribers may purchase copies of individual Reports at £1 each. Earlier Reports are at lower prices.

For further information and a catalogue of publications, please write to the Report Editor at Coombe Lodge, Blagdon, Bristol, BS18 6RG, or telephone Blagdon (0761) 62503.

Previous Publications

Vol 7 No 1	The Organisation of Engineering Education
2	The Management of College Departments – Change and Development
3	Developments in Technician Education (TEC)
4	Staff Development in Scottish Colleges
5	Polytechnic Developments
6	The F.E. College and the Community
7	Developments in Art Education
8	Colleges of Education and Reorganisation
9	Social Work Education
10	Schools and Colleges 16-19 The Education of the Less Able Student
11	The Dip. H.E.
12	College Government